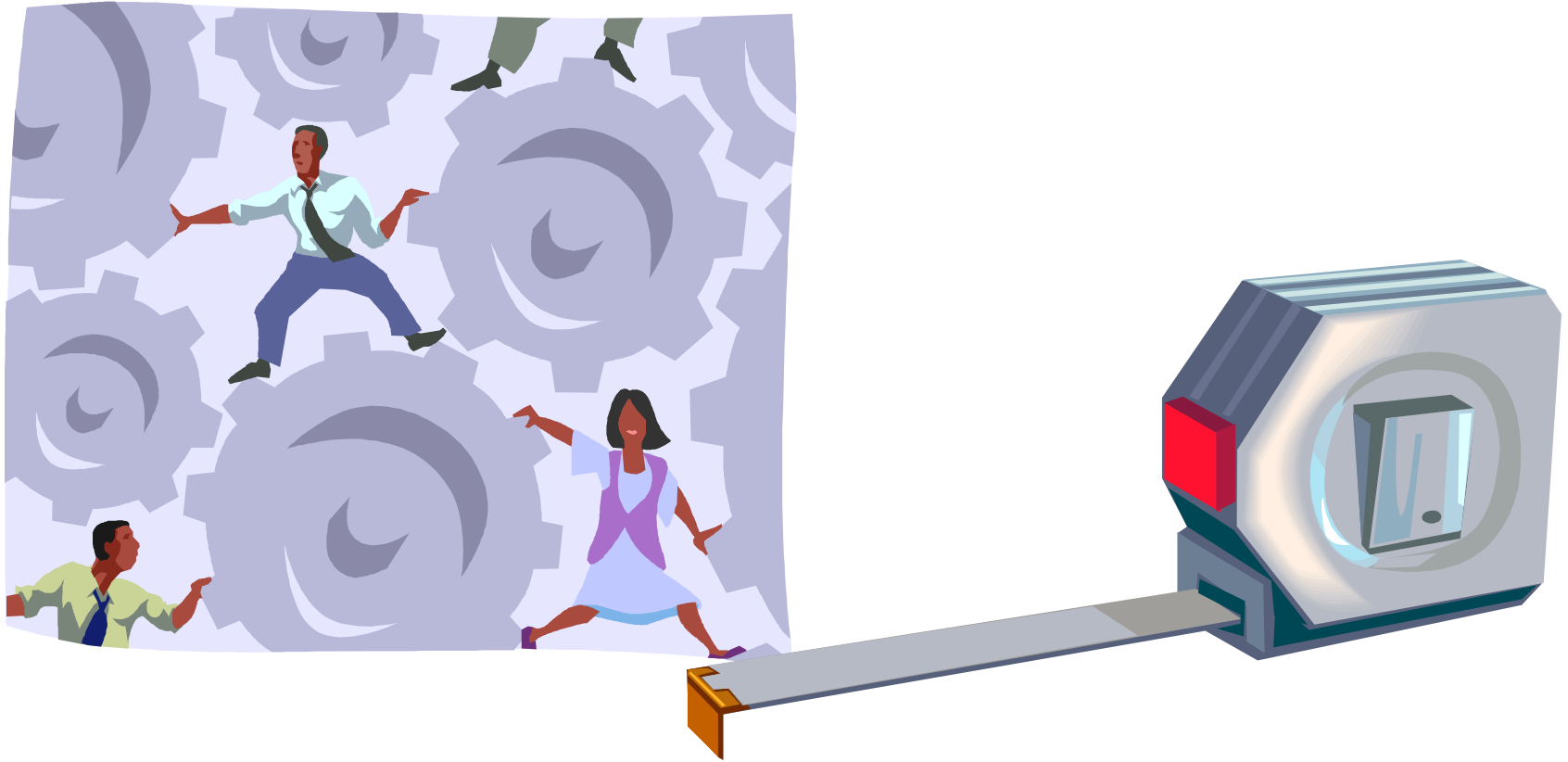


Current Metrics Initiatives



Smith-Jentsch, K. A. & Burke, C. S. (2007). Current metrics initiatives. *Presentation to the Office of Naval Research Collaboration and Knowledge Interoperability Program, Arlington, VA, August 9th, 2007.*



| Report Documentation Page | | | | Form Approved OMB No. 0704-0188 | |
|--|------------------------------------|-------------------------------------|--|---|------------------------------------|
| Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. | | | | | |
| 1. REPORT DATE 2007 | | 2. REPORT TYPE N/A | | 3. DATES COVERED - | |
| 4. TITLE AND SUBTITLE Current Metrics Initiatives | | | | 5a. CONTRACT NUMBER | |
| | | | | 5b. GRANT NUMBER | |
| | | | | 5c. PROGRAM ELEMENT NUMBER | |
| 6. AUTHOR(S) | | | | 5d. PROJECT NUMBER | |
| | | | | 5e. TASK NUMBER | |
| | | | | 5f. WORK UNIT NUMBER | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Central Florida | | | | 8. PERFORMING ORGANIZATION REPORT NUMBER | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | | | 10. SPONSOR/MONITOR'S ACRONYM(S) | |
| | | | | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) | |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited | | | | | |
| 13. SUPPLEMENTARY NOTES The original document contains color images. | | | | | |
| 14. ABSTRACT | | | | | |
| 15. SUBJECT TERMS | | | | | |
| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT UU | 18. NUMBER OF PAGES 60 | 19a. NAME OF RESPONSIBLE PERSON |
| a. REPORT unclassified | b. ABSTRACT unclassified | c. THIS PAGE unclassified | | | |

Primary Metrics Researchers

- Kim Smith-Jentsch – UCF
- Shawn Burke – UCF
- Valerie Sims – UCF
- Nancy Cooke - ASU



Experimentation

- Refine a nomological network regarding macrocognition
 - How many distinct components?
 - How are they related?
 - How are they affected by contextual variables?



Overarching Objective

To identify a set of metrics that are unobtrusive, construct valid, incrementally predictive, and sensitive to our manipulations for use in experimentally-testing our hypotheses.



Experimental Tasks

- Validity of metrics should generalize across tasks that share substantive features

- Two task environments have been developed and pilot-tested
 - NEO scenario
 - ER simulation





Our Philosophy Regarding Metrics



Process v. Product at Team or Individual Level of Analysis

| | Product | Process |
|------------|---|---|
| Individual | <ul style="list-style-type: none">• Concept maps | <ul style="list-style-type: none">• Point of regard at key events |
| Team | <ul style="list-style-type: none">• Degree of overlap | <ul style="list-style-type: none">• Complementary point of regard at key events |

Components of a Metric

- Content: for example, terms, behaviors, physiological reactions
- Source: for example, participant, peers, outside observers, equipment
- Method: for example, Likert-type scale, checklist, card-sorting task, pairwise comparisons
- Scoring/indexing: for example, percentage, mean, sum, correlation to expert, distance from expert



Content



Content

- Cognitive: lots of construct confusion
- Physiological: very little done in team arena, however some preliminary evidence that teams with similar reactions perform better
- Attitudinal: levels of analysis issues
- Behavioral: problems with discriminant validity



Methods

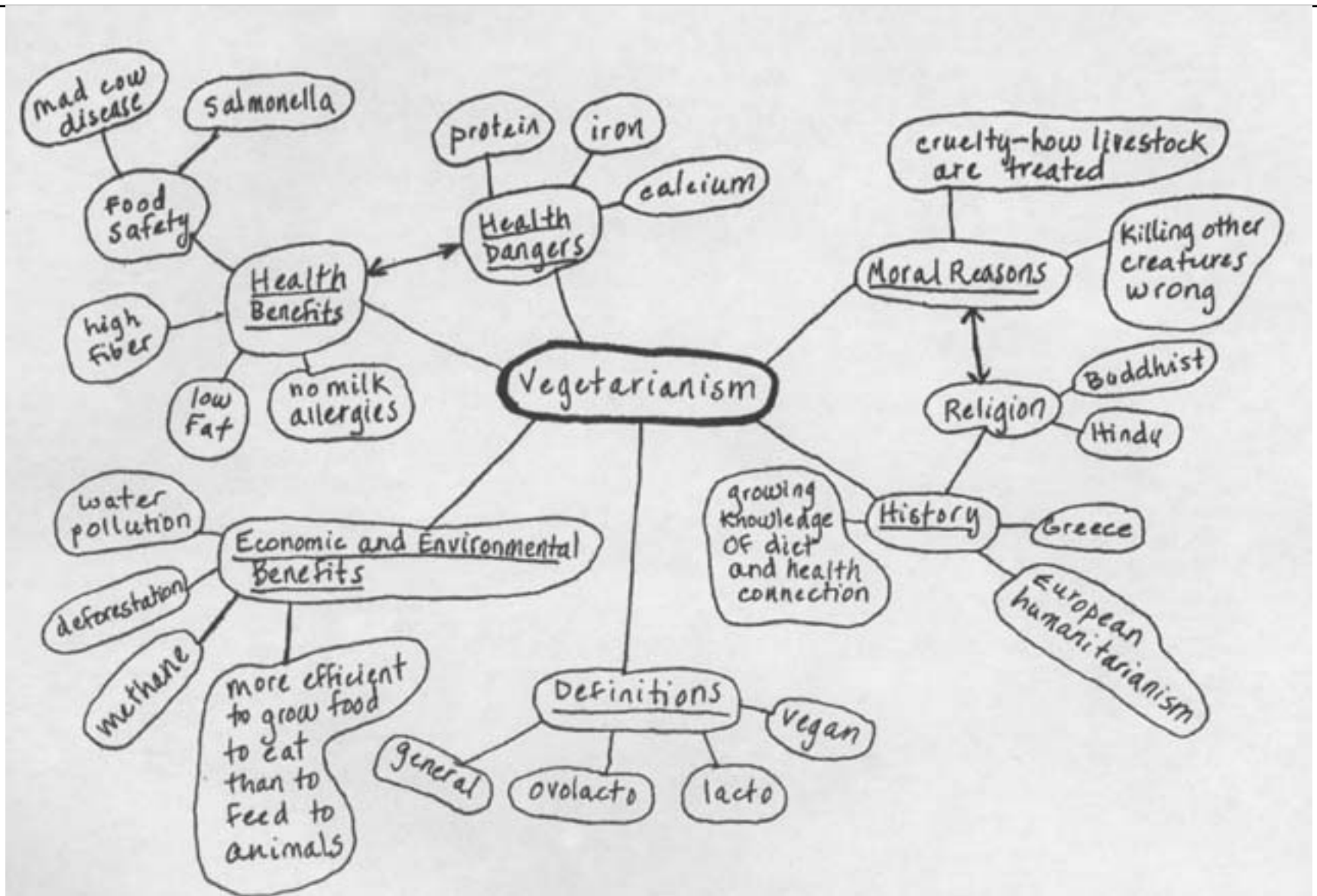


Major Methods for Measuring Cognitive Team Constructs

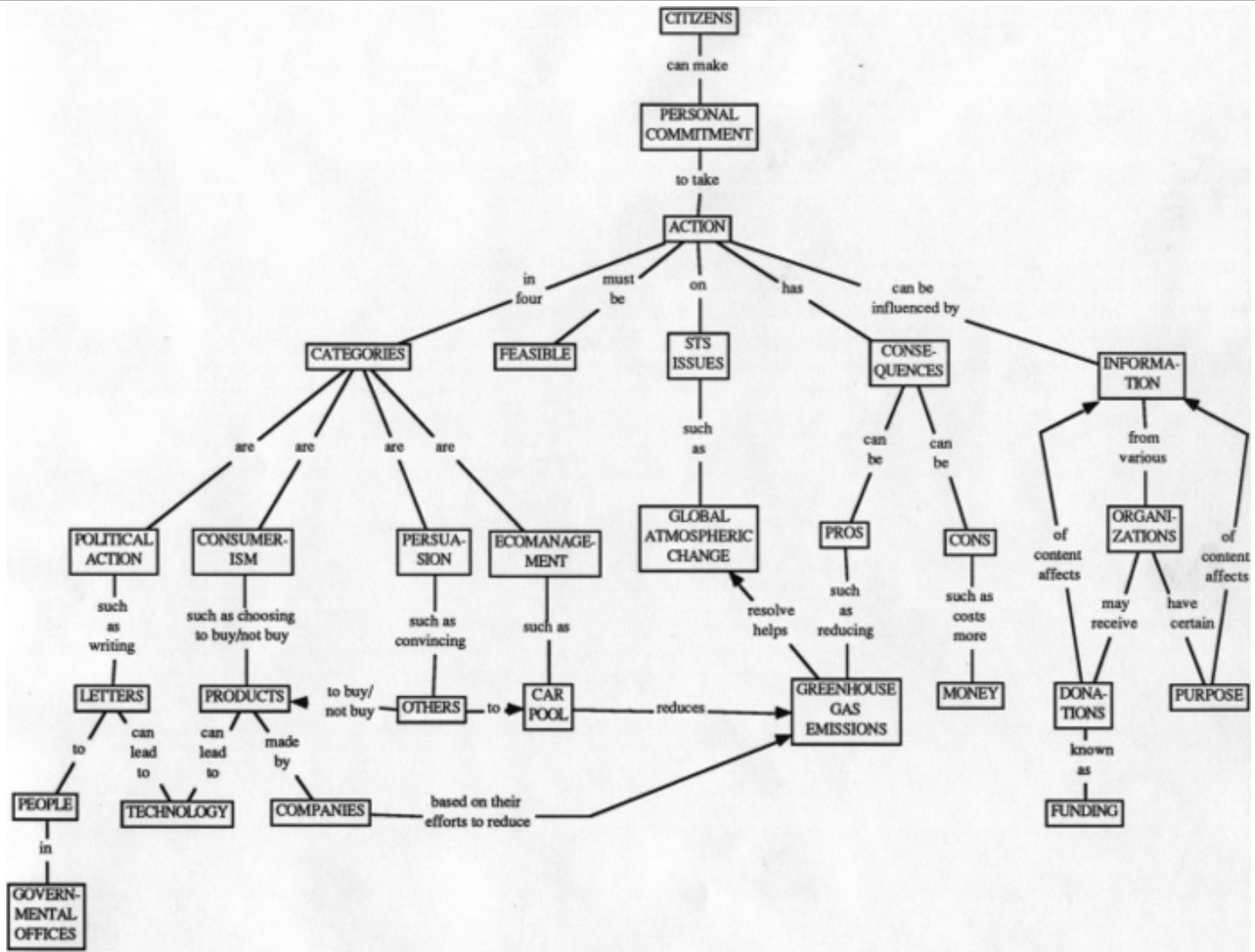
- ❑ Concept maps
- ❑ Pair-wise comparison ratings
- ❑ Card sorting
- ❑ Vignette-based ratings
- ❑ Think-aloud protocols
- ❑ Probes during task performance
- ❑ NEED TRIANGULATION



Concept-Map



Concept-Map



Card Sorting Task

Information
Exchange

At the beginning
the manager told
staff how many of
the kitchen had
reservations the
there were any
comin

**The hostess offered to take
drink orders for the table she
had just seated since the
server had his hands full with
a large party.**

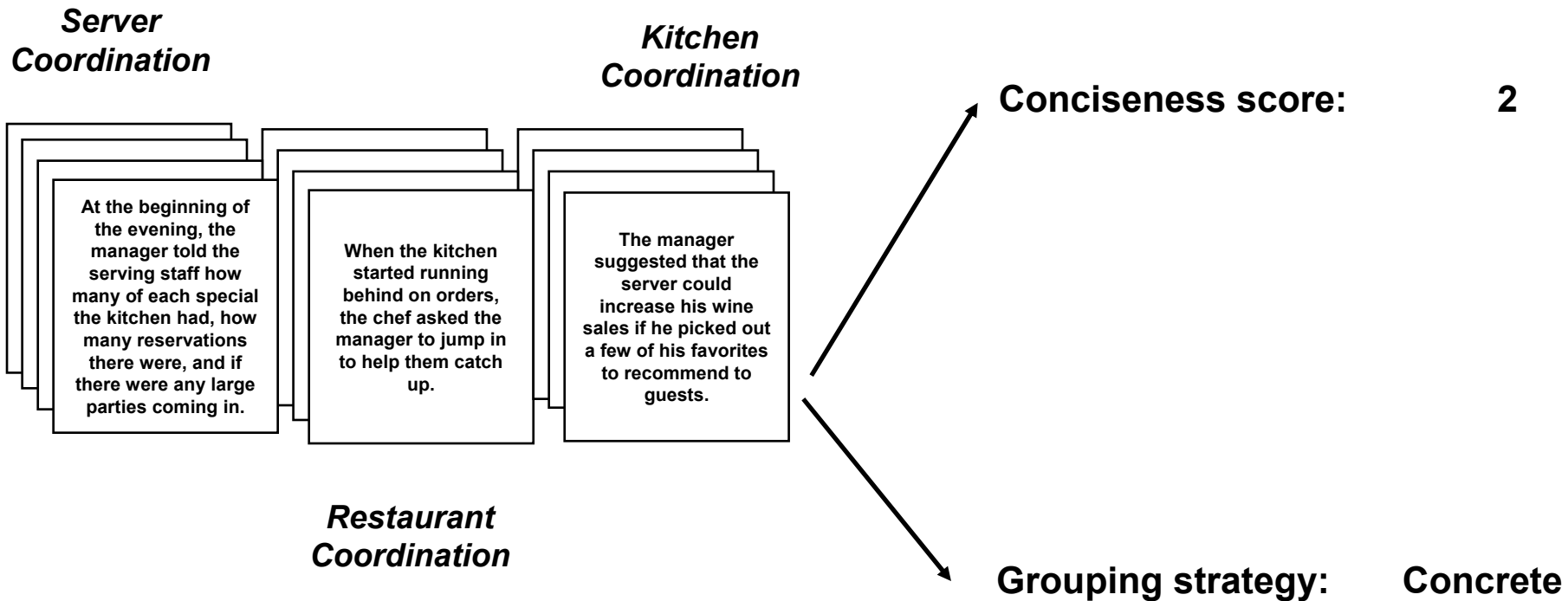
Initiative/
Leadership

er suggested that the
ld increase his wine
icked out a few of his
to recommend to
guests.

Behavior



Conciseness & Grouping Strategy



Vignette-Based

Teammate A appears to you to be overloaded and in need to assistance. If you were to offer assistance to teammate A how likely would he/she be to:

Defensively refuse your assistance:

Politely refuse your assistance:

Gratefully accept your assistance:

Grudgingly accept your assistance:

Embarrassingly accept your assistance:

Highly unlikely

Highly likely

1-----2-----3-----4-----5-----6

Relatedness Ratings

Participant's Ratings SME's Ratings

To what degree are the following goals related and in which direction?



Major Methods for Measuring Behavioral/communication-related Team Constructs

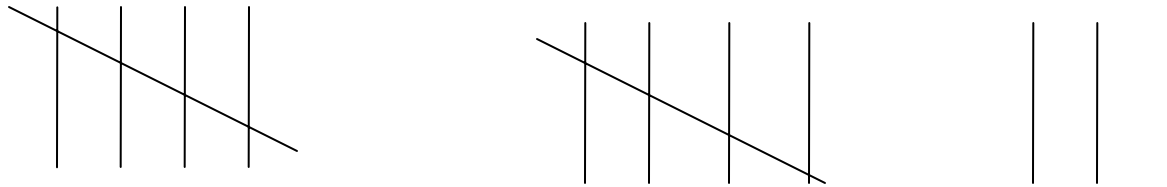
- Likert scales
- Frequency counts
- Behaviorally anchored rating scales
- Checklists
- Communication flow/sequencing

Each have strengths and weaknesses – construct and context must be considered when determining which to use!



Frequency Counts

Number of times that backup was demonstrated:



Likert-type Scale Backup Behavior Was:

Highly Ineffective Highly Effective
1-----2-----3-----4-----5-----6

Behaviorally Anchored Rating Scale (BARS)

| | |
|-----|---|
| 5 | Typically, in this organisation, one would expect mechanisms to have been introduced for the express purpose of cascading information systematically from top to bottom of the organisational hierarchy. |
| 4.5 | Typically, in this organisation, one would expect a management information provision unit to have been established which is constantly consulting information users on their present and future needs. |
| 4 | Typically, in this organisation, one would expect minutes of governing body meeting (e.g. Board of Governors/Council/Academic Board) to be made available to all staff, and actively circulated to those who need them. |
| 3.5 | Typically, in this organisation, one would expect information bulletins from management, and meetings, to focus primarily on developments that have already taken place, as opposed to developments in the pipeline. |
| 3 | Typically, in this organisation, one would expect there to be both formal and informal channels for information but information provision to be not always timely. |
| 2.5 | Typically, in this organisation, one would expect there to be widespread use of the 'confidential' and 'restricted' stamp on documents and reports. |
| 2 | Typically, in this organisation, one would expect information produced centrally to be consigned to the waste bin frequently by recipients because it is thought to serve no useful purpose. |
| 1.5 | Typically, in this organisation, one would expect information provision to be 'ad hoc' in the sense of being provided when requested if one happens to know that it is available and the relevant party to contact. |
| 1 | Typically, in this organisation, one would expect little or no publicity to be given to major developments such as the setting up of a new unit or the introduction of a new facility. |

Checklists

Assisted AAWC in identifying tracks at first wave.

Assisted AAWC in identifying tracks at second wave.

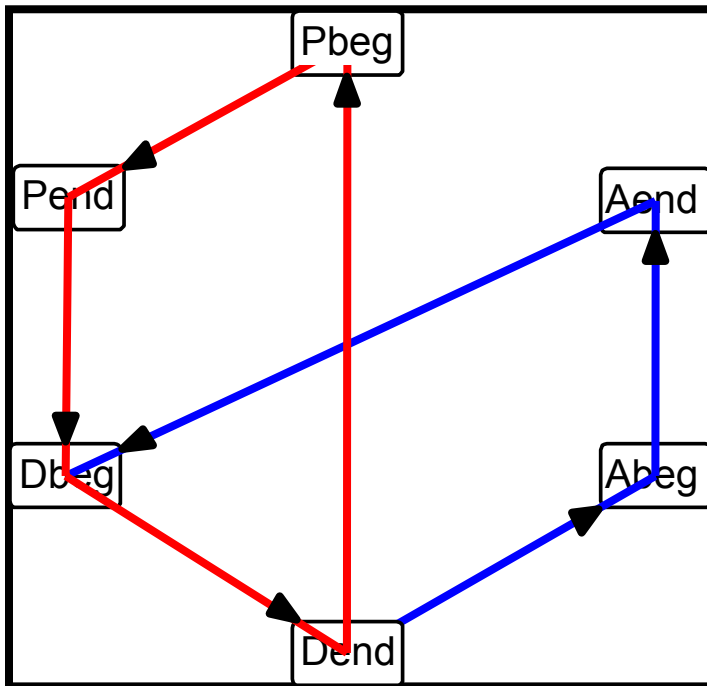
Assisted AAWC in identifying tracks at third wave.

Assisted AAWC in identifying tracks at fourth wave.

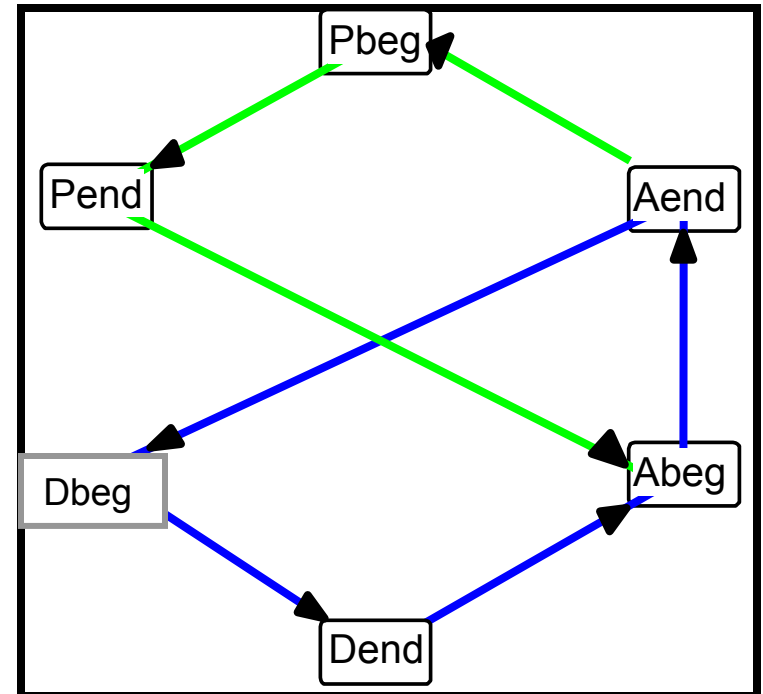
| YES | NO |
|-----|----|
| X | |
| | X |
| | X |
| X | |

Flow Patterns Change in Real-time

Before Conflict



After Conflict



Major Methods for Measuring Physiological Team Constructs

- Synchronous:
 - Eye scan/fixation
 - Pupil dilation
 - Increases in heart rate/blood pressure
 - Vocal intensity, pitch



Fixation

Saccade

Hospital Rules

Rule 1: Patients in the ER are not seen by a doctor on a first-come/first served basis: The triage nurse determines their order on the basis of need.

Rule 2: Patient information is confidential. Each patient has the right to decide:

- Who may see him or her
- Who may have information about their condition
- Whether you may reveal that they are even registered as a patient in the hospital

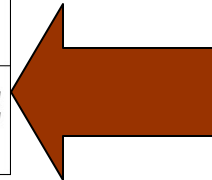
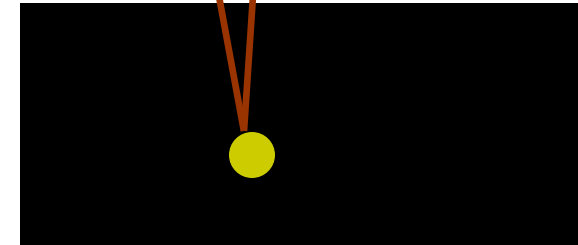
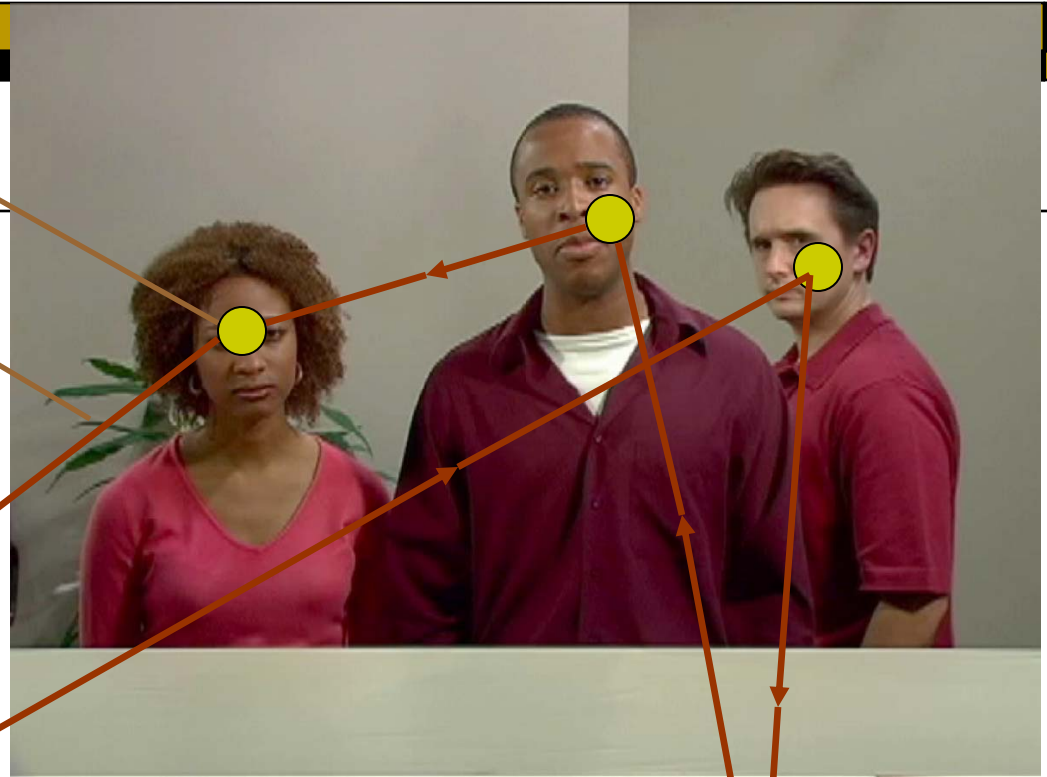
Rule 3: Doctors have the right to deny access to a patient if it is in the best interest of the patient's medical condition.

Rule 4: You will be fired if you lie verbally or in writing.

Rule 5: No racial or sexual harassing comments.

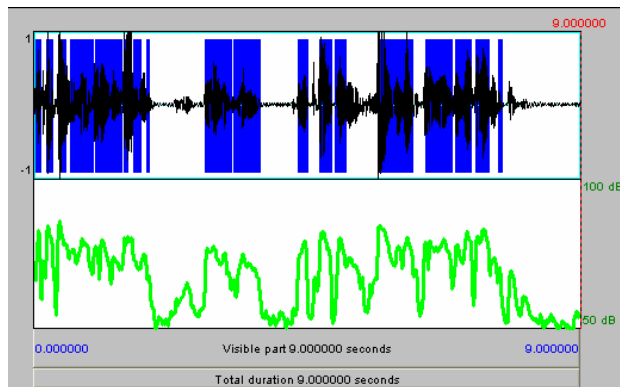
Patient Record

| <u>Patient Name</u> | <u>Gender</u> | <u>Consent</u> | <u>Medical Condition</u> | <u>Notes</u> |
|---------------------|---------------|--|-------------------------------------|---|
| Michael Rayfield | Male | Doctors day no visitors | Severe trauma from auto accident | Currently in the intensive care unit |
| Jane Doe | Female | Non-publicity patient | Complications during miscarriage | Real name: Kayla Johnson |
| Carmen Diaz | Female | Only grandfather, Manny Diaz, may see Carmen | Several broken bones and concussion | Mother suspected of child abuse. She's <i>not</i> allowed to see the child. |

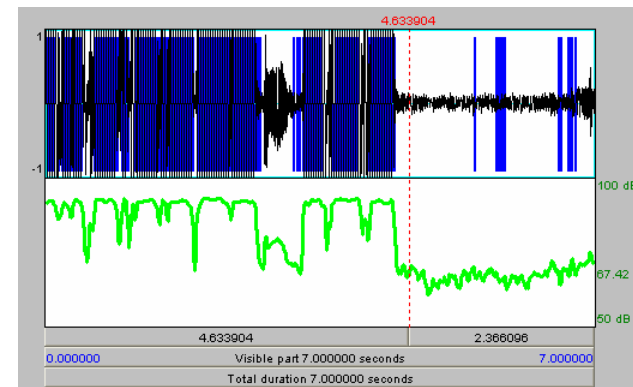


- Intensity analysis indicates the degree of aggressive or assertive tendencies between team members.

Participant A

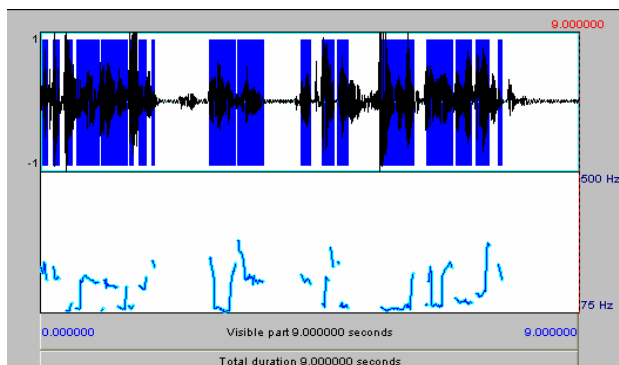


Participant B (More Aggressive)

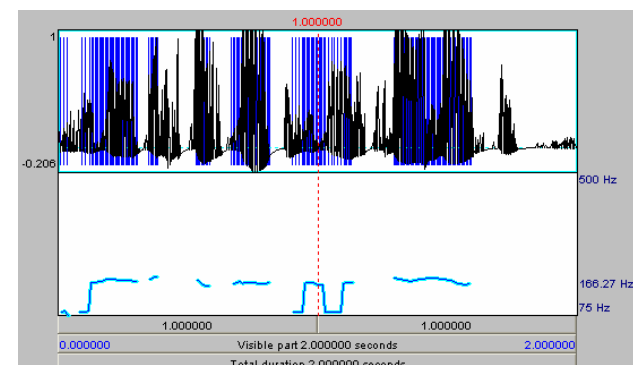


- Pitch analysis indicates the degree of emotion on the person's voice in a stressful scenario.

Participant A (More emotional)



Participant B



Autonomic Measures and Findings

| Measure | Related Construct | Previous Findings | Reference |
|--------------------------------|--|--|---------------------------|
| Heart Rate (HR) | Self monitoring | HR activity during anticipation of socially threatening situation (public speech) was negatively related to self monitoring. | Hofman, 2006 |
| | Self-efficacy | Positively associated with hr and unrelated to self reported arousal. | Gellatly & Meyer, 1992 |
| | Personal goal level | Positively associated with hr and unrelated to self reported arousal. | Gellatly & Meyer, 1992 |
| | Goal difficulty | Positively associated with HR. | Gellatly & Meyer, 1992 |
| | Performance | HR positively associated with task performance but unrelated to self reported arousal. | Gellatly & Meyer, 1992 |
| | Team performance | Team change in HR similarity positively predicted overall performance and errors made on a team coordination task involving robotic operation. | Henning et al., 2001 |
| Blood pressure | Task difficulty Social Presence | Blood pressure higher when being observed compared to no observation while a performing difficult task. This was not found for an easy task. | Henning et al., 2001 |
| Galvanised Skin Response (GSR) | Self Monitoring Social Threat | GSR activity during anticipation of socially threatening situation (public speech) was negatively related to self monitoring. | Hofman, 2006 |
| | Learning | GSR strength during training predicted learning in support of Damasio's Somatic Hypothesis. | Carter & Pasqualini, 2004 |
| Cortisol | Stress/Anxiety | Cortisol sensitive to stress and anxiety. | Schlotz et al., 2006 |



Criteria for Metrics



Criteria for Metrics

- ❑ Construct validity
- ❑ Incremental predictive validity
- ❑ Sensitivity
- ❑ Unobtrusiveness



Construct Validity

- ❑ To what degree does a metric capture the construct of interest?
- ❑ Convergent validity: Should correlate highly with other metrics of the same construct - TRIANGULATION
- ❑ Discriminant validity: Should not correlate highly with metrics of theoretically distinct constructs.

Sensitivity

- ❑ To what degree does the metric detect differences among participants related to other variables/manipulations of interest?
- ❑ Floor and ceiling effects
- ❑ Quantity versus quality
- ❑ Ordinal versus interval or ratio scales



Unobtrusiveness

- To what degree does administration of the metric itself influence participants
- A potential source of contamination
- “Testing” threat to validity
- Can be assessed by comparing participants who were administered the metric (e.g., eyetracking) to those who were not on the dependent variables of interest



Incremental Predictive Validity

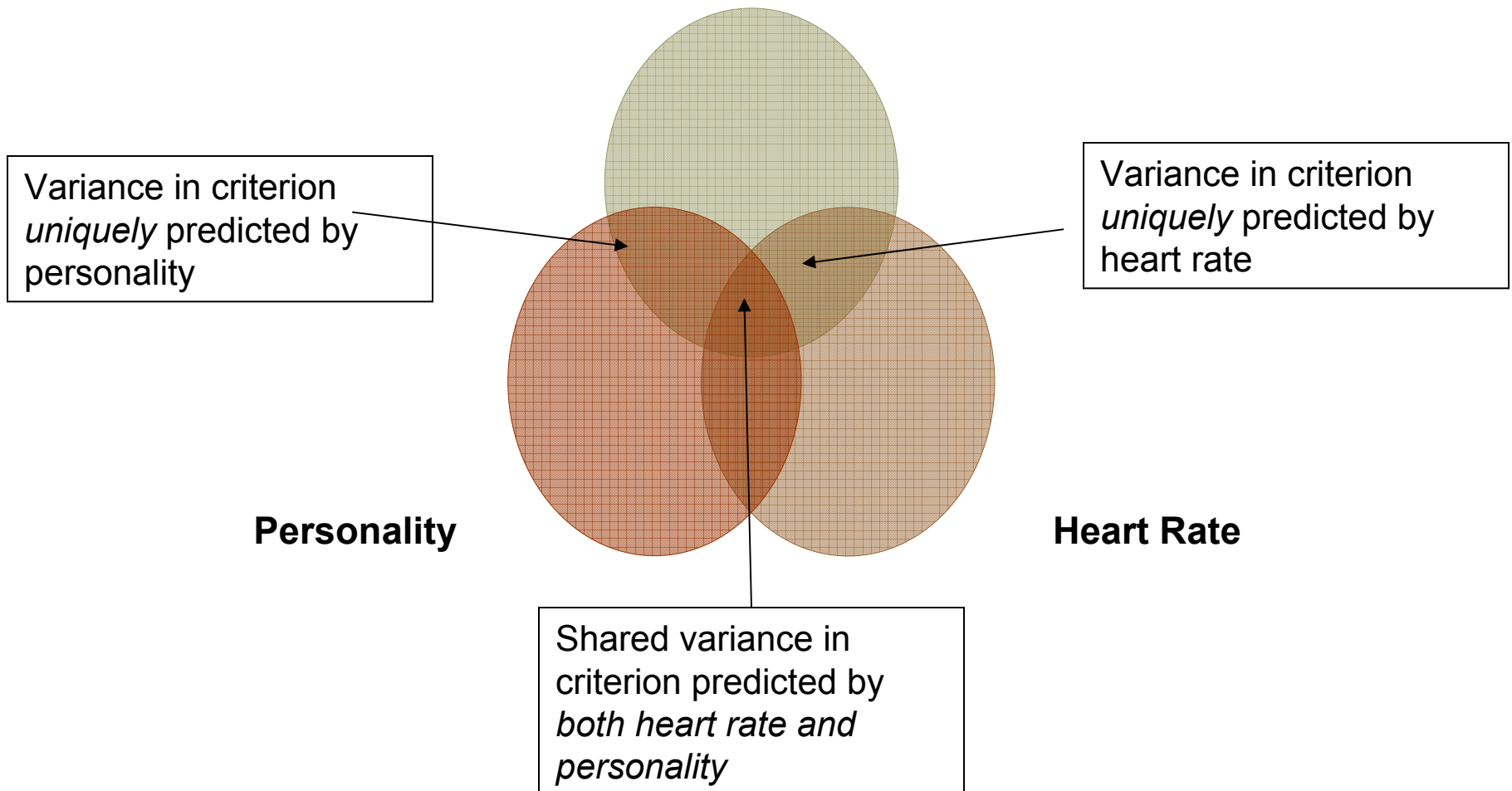
- To what degree do multiple measures of macro-cognition contribute uniquely to the prediction of team performance (explain unique variability)?

- Can be assessed using multiple regression analyses
 - Multiple metrics entered as predictors of a DV
 - Each metric significant beta when all considered together?
 - Significant change in R-squared when new metric added?



Incremental Predictive Validity

Sharing Unique Information





Approach to the Study of Metrics



Emerging Lessons


- ❑ Experimental/theoretical constructs lack clarity
- ❑ Primary focus on emergent states; lack of attention to development of team cognition or corresponding cognitive processes
- ❑ Incremental validity often not tested
- ❑ Content and method often confounded
- ❑ Over reliance on self report and single method
- ❑ Method (Messmer-Magnus & DeChurch, 2007), and indexing choices (Smith-Jentsch, Mathieu, & Kraiger, 2005) make a difference



Major Designs

- Meta-analysis (in-progress)
- Predictive validity studies (first in-progress)
- Metric “bake-off’s”
- Experimental tests of obtrusiveness
- Manipulation of task characteristics





Meta-Analysis of Prior Studies Measuring Components of Macro cognition



Cognitive Team-related Constructs

| CKI Model | Schema Theory | Transactive Memory | Mental Model Theory | Situation Awareness Theory |
|---|----------------------|---|-----------------------------------|-----------------------------------|
| Recognizing Expertise | Teammate Schema | Teammate knowledge consensus & accuracy | Mental models of teammates | |
| | | | Mental models of team interaction | |
| Individual Knowledge Development | | Knowledge stock | Task & Equipment mental models | |
| Team Problem Model Pattern Recognition | | | | Team Situation Awareness |

Quantitative Database

| | Content | | | Elicitation | | Aggregation | Potential Moderators & Mediators |
|-----------------------|---|---|---|--|---|---|---|
| Method | Construct Validity | Knowledge Type | Emergent Outcome(s) Focused On | Method | Source | Aggregation Method | Task/Team Properties |
| Mathieu et al. (2000) | <ul style="list-style-type: none"> Task Team | <ul style="list-style-type: none"> Declarative | <ul style="list-style-type: none"> Sharedness | <ul style="list-style-type: none"> Pairwise ratings | <ul style="list-style-type: none"> Self-report | <ul style="list-style-type: none"> UCINET- QAP correlation (task had two members) | <ul style="list-style-type: none"> 2 person teams Flight combat sim. Novice, lab |
| Espevik et al. (2006) | <ul style="list-style-type: none"> Equipment Team Team interaction | <ul style="list-style-type: none"> Declarative | <ul style="list-style-type: none"> IPK | <ul style="list-style-type: none"> Questionnaire | <ul style="list-style-type: none"> Self-report | <ul style="list-style-type: none"> N/A | <ul style="list-style-type: none"> 6 person teams Tactical submarine sim. Active duty officers |
| Marks et al. (2000) | <ul style="list-style-type: none"> Team interaction | <ul style="list-style-type: none"> Procedural | <ul style="list-style-type: none"> 3-way overlap | <ul style="list-style-type: none"> Pairwise ratings | <ul style="list-style-type: none"> Self-report | <ul style="list-style-type: none"> Pathfinder C index for each pair of team members averaged | <ul style="list-style-type: none"> 3 person teams Apache helicopter sim. Novice, lab |
| | <ul style="list-style-type: none"> Team interaction | <ul style="list-style-type: none"> Procedural | <ul style="list-style-type: none"> 3-way overlap | <ul style="list-style-type: none"> Concept mapping | <ul style="list-style-type: none"> Self-report | <ul style="list-style-type: none"> % of shared concepts placed identically on map | <ul style="list-style-type: none"> 3 person teams Tank sim. Novice, lab |



Predicting Effective Team Problem Solving Using Metrics of Macrocognition: How Many Distinct Constructs Exist and What are Their Relations?





Relations Among Macrocognitive Constructs

- Initial experiment designed and approved by UCF Human Subjects Review Board
- Piloted using 40 participants this summer
- Scheduled to begin actual study in two weeks



Secondary Task Environment



Experimental Tasks

□ Commonalities

- Non-combatant evacuation scenarios
- Require team problem solving processes
- Must resolve ambiguity
- Product of the team's performance is a plan
- Both allow for face-to-face or distributed communication

□ Differences

- Team size
- Hierarchical position in multi-team system
- Degree to which task involves social cues
- Distance from the disturbance



Hospital Emergency Room Simulation



Team Roles

Customer service personnel
(participants)



Remote Hospital Personnel
(simulated characters)

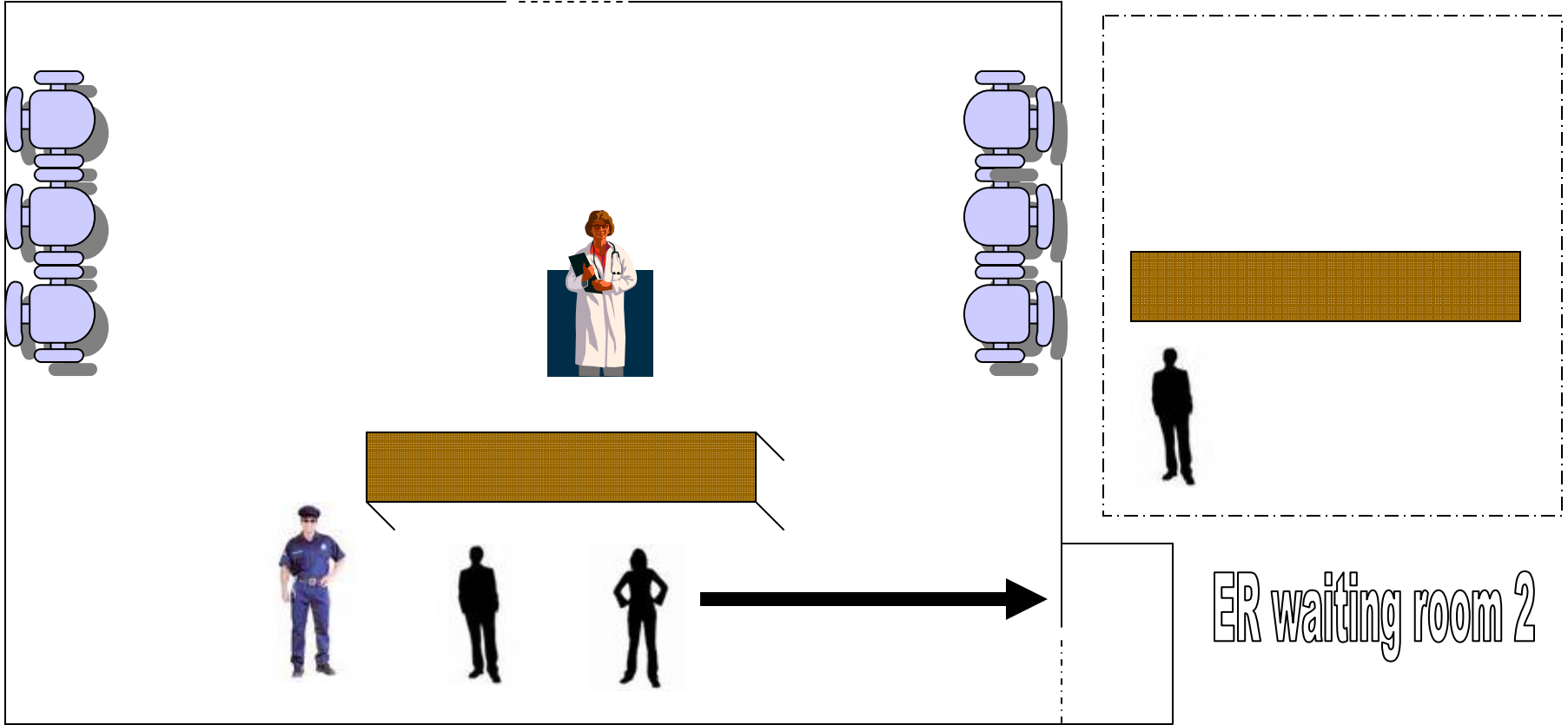


ER doctor/nurses
(live confederate)



Police officer
(live confederate)

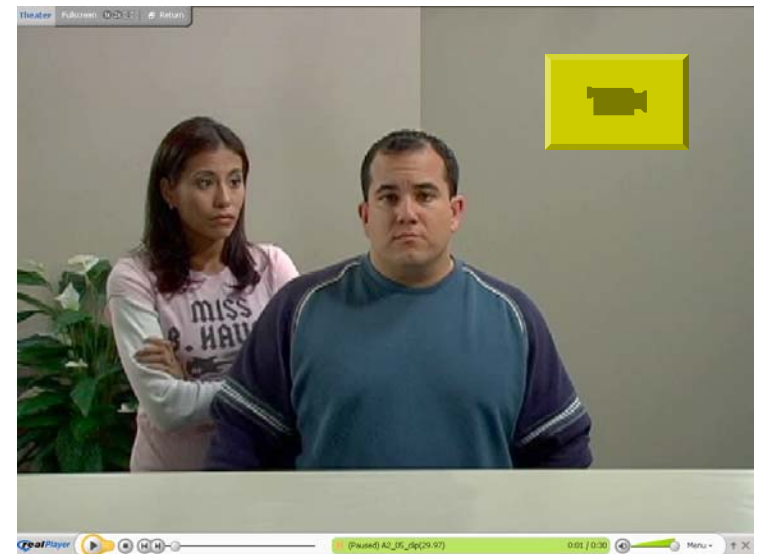




ER waiting room 1

ER waiting room 2

Scripted Events



Shared and Unshared Data

- Participants share a common view of ER waiting room
- Participants also have access to independent data screens
- Must perform routine planning tasks while servicing customers prior to the conflict event.



Problem Solving Element

- After interacting with video-based and live confederates, participants are faced with a hostage situation
- Must aid police in determining how many hostages, who they are, who is involved in the disturbance and what it is about.
- Clues are revealed as situation unfolds



Macrocognitive Constructs to be Measured in Initial Study

- Team problem models
(a.k.a. situation awareness)
 - Methods: Paired comparison ratings, card-sorting

- Sharing of unique knowledge
(a.k.a information exchange)
 - Methods: Eye-tracking, communication analysis



Relatedness Ratings

Participant's Ratings
SME's Ratings



Card Sorting Task

Other Patients in Lobby



Angry People



Macrocognitive Constructs to be Measured in Initial Study

- Team problem models
(a.k.a. situation awareness)
 - Method: Paired comparison ratings, card-sort

- Sharing of unique knowledge
(a.k.a information exchange)
 - Methods: Eye-tracking, communication analysis



Fixation

Saccade

Hospital Rules

Rule 1: Patients in the ER are not seen by a doctor on a first-come/first served basis: The triage nurse determines their order on the basis of need.

Rule 2: Patient information is confidential. Each patient has the right to decide:

- Who may see him or her
- Who may have information about their condition
- Whether you may reveal that they are even registered as a patient in the hospital

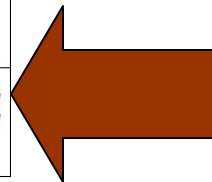
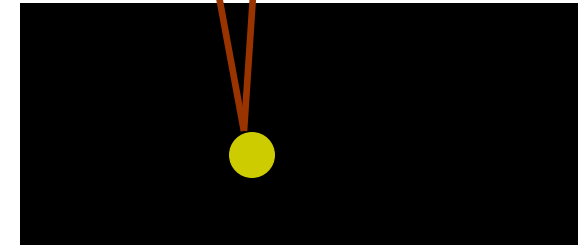
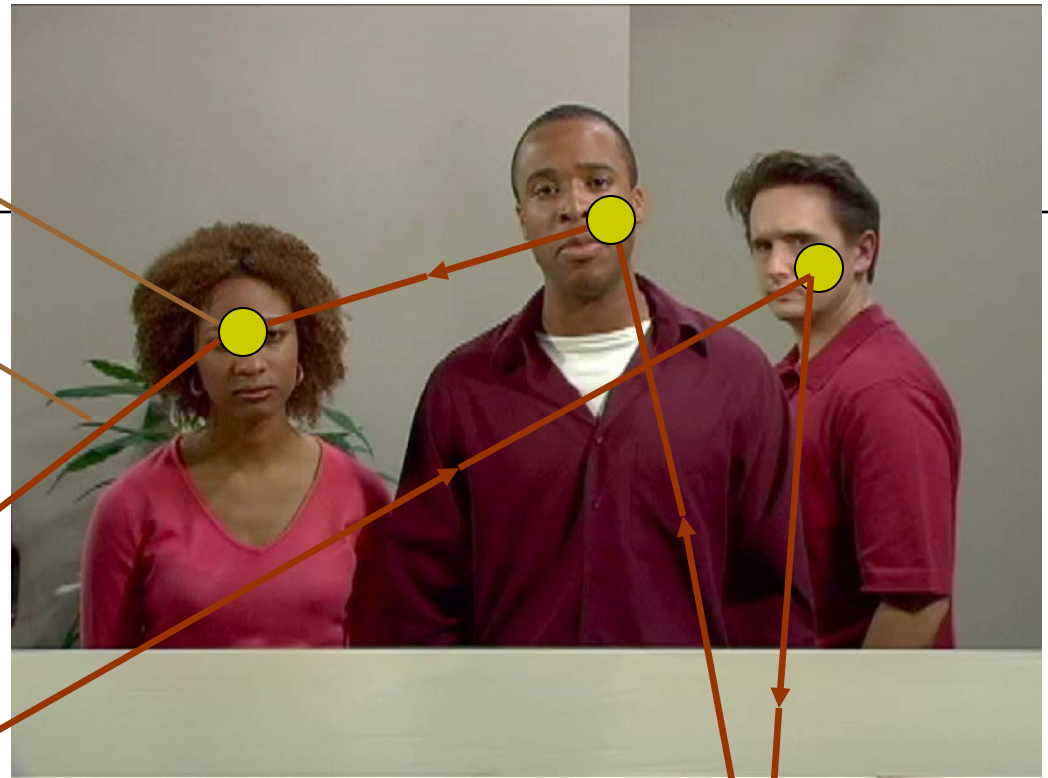
Rule 3: Doctors have the right to deny access to a patient if it is in the best interest of the patient's medical condition.

Rule 4: You will be fired if you lie verbally or in writing.

Rule 5: No racial or sexual harassing comments.

Patient Record

| <u>Patient Name</u> | <u>Gender</u> | <u>Consent</u> | <u>Medical Condition</u> | <u>Notes</u> |
|---------------------|---------------|--|-------------------------------------|---|
| Michael Rayfield | Male | Doctors day no visitors | Severe trauma from auto accident | Currently in the intensive care unit |
| Jane Doe | Female | Non-publicity patient | Complications during miscarriage | Real name: Kayla Johnson |
| Carmen Diaz | Female | Only grandfather, Manny Diaz, may see Carmen | Several broken bones and concussion | Mother suspected of child abuse. She's <i>not</i> allowed to see the child. |



Macrocognitive Constructs to be Measured

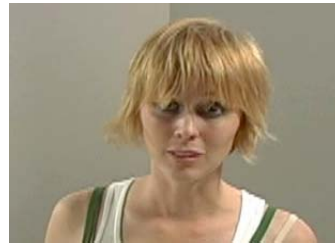
- Stress
 - Methods: Heart rate, self-report

- Communication delivery
 - Methods: Pitch, intensity, Likert-type scales

- Problem solving accuracy
 - Method: Checklist



Who is still in the lobby, who is likely involved in the disturbance, what is it about?



Pilot Data from Secondary Task

- Physiological metrics correlate with communication analysis
 - Heart rate and blood pressure predict information sharing
 - Vocal intensity predicts communication delivery
- Eye tracking data accurate enough to reliably distinguish objects the following distances apart from one another:
 - 7 inches at 8 feet away (across the room)
 - 3 inches at 2 feet away (on a monitor)
- Eye tracking equipment does not bother participants



Comfort Level of Eye-tracking goggles
1 (not at all) - 5 (a great deal)

| Questions | Means | Standard Deviations |
|--|--------------|----------------------------|
| To what degree did the eye tracking equipment obstruct your view during simulation task? | 1.50 | .548 |
| To what degree did the eye tracking equipment obstruct your view during the card sort task? | 1.33 | .516 |
| To what degree did you have to adjust to the eye tracker? | 1.50 | 1.225 |
| To what degree did the eye tracker caused discomfort or irritation to your eyes? | 2.00 | 1.095 |
| To what degree did the weight of the eye tracker equipment caused you discomfort? | 1.67 | 1.211 |
| To what degree did you feel intimidated having to wear the eye tracking equipment? | 2.33 | 1.633 |
| To what degree did the eye tracker equipment affect your ability to use the mouse during your tasks? | 1 | 0 |
| To what degree did the eye tracker equipment affect your ability to use the keyboard during your tasks? | 1.50 | .548 |
| To what degree did the eye tracker equipment make you feel awkward or uncomfortable while interacting with co-workers, patients & supervisors in the simulation? | 2.00 | 1.265 |
| To what degree did the eye tracker equipment distract your overall performance? | 1.50 | .837 |
| Overall, how uncomfortable was it to wear the eye tracking goggles? | 2.17 | .983 |

Summary

- Seeking to uncover a measurable and parsimonious model of macrocognition
- Searching for converging results across research methodologies
- Triangulation of measurement strategies

